

董 军/著

人工智能哲学

探索人工智能本质 揭示思维科学内涵



科学出版社

人工智能哲学

董 军/著

科学出版社

北京

图书在版编目(CIP)数据

人工智能哲学 / 董军著. —北京: 科学出版社, 2011. 9

ISBN 978-7-03-032372-9

I. ①人… II. ①董… III. ①人工智能—技术哲学 IV. ①TP18-02

中国版本图书馆 CIP 数据核字 (2011) 第 191011 号

责任编辑: 侯俊琳 郭勇斌 卜 新 / 责任校对: 李 影

责任印制: 赵德静 / 封面设计: 黄华斌

编辑部电话: 010-64035853

E-mail: houjunlin@mail.sciencep.com

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2011 年 10 月第 一 版 开本: B5 (720 × 1000)

2011 年 10 月第一次印刷 印张: 12 1/2

印数: 1—3 000 字数: 189 000

定价: 35.00 元

(如有印装质量问题, 我社负责调换)

前 言

在现实生活中，哲学给人既高深又平常的双重印象，在相当长的一段时间里，我国的学生在其学生生涯的一开始就接受马克思主义哲学教育，所谓“从小粗知一点马克思主义”，包括辩证唯物主义和历史唯物主义，涉及世界观、方法论等。这些理论是实践的产物、历史的总结，是合乎规律的认识论。但是，在这样的教育安排下培养出来的教师、工程师、公务员以至教授、科学家大都没有认识到哲学的意义和价值，几乎一直认为哲学是教条，甚至与政治混同，从这种实际情况看，哲学教育要实现给人以“智慧”的目的，需要在方法、内容、目标、途径和实践等环节进一步反思。

放眼世界，哲学被纳入教育的基本内容不是孤立的情形。哲学的概念源自古希腊。哲学教育的差别在于以什么样的形式和手段传授哪些知识和思想。在学校里，有学生肯定哲学老师授课精彩的情形，这表明教学方式的选择和课程内容的安排是关键。但哲学授课的目的不仅仅是要同学们说“有趣”，更是要教给他们看待日常问题的方法、关于世界的基本观点和面对人生的洞达态度。

不同程度的学生需要不同层面的知识与技能。本科生的主要任务是专业基础知识的掌握和学习方法的培养；硕士研究生要拓宽眼界，经历科研；博士研究生要能够思考问题，分析问题。无论是“提出问题比解决问题更重要”，还是“正确的问题是成功的一半”，都需要科技工作者以良好的方法、宽广的视野在前人工作的基础上“开题”，所谓“创新”因此而有可能。

20 世纪 80 年代初，伴随着改革开放的脚步，西方各种思潮开始涌入中国内地。从弗洛伊德《梦的解析》到卡西尔的《人论》等，上海译文出版社陆续出了《二十世纪西方哲学译丛》，有的书籍销售量以十万计。1986 年，我

在书店偶尔购得钱学森先生主编的《关于思维科学》。1988年，我第一次去北京出差时，秋季书市还没结束，购买了马克思的《资本论》等。我那时工作在国营企业，上班空闲时通读了父亲收藏的《马克思恩格斯选集》以及自购的《资本论》等。据同事在很多年以后说起，那时，清华大学毕业的总工程师等对我读这类与本职工作“无关的”书籍颇有微词（尽管我在那里曾以最高票当选过年度“先进工作者”）。1991年，我攻读硕士学位入学前，读了在南京出差时偶尔购得的蔡自兴、徐光佑教授的《人工智能及其应用》（第一版），这本书带我进入人工智能之门，而《关于思维科学》则犹如“经典”，对若干篇的内容读过多遍。

钱学森先生认为，我们有马克思主义的指导，应该比外国人做得更好。1977年，当时年轻数学家杨乐、张广厚研究员在函数论方面工作出色，他们谈到毛泽东的《实践论》、《矛盾论》对其工作的指导意义。我听何积丰院士不止一次谈到《实践论》、《矛盾论》，讲过认识的螺旋式上升理论是逐渐才能被理解的。杰出书法家、书学家、书法教育家祝嘉先生说过，能仔细读一下《资本论》等，可使心胸开朗，眼界放大，必大有好处。

西方的哲学思想，有的晦涩如象牙塔，有的纯粹为了建某个体系，有的面向生活，有的实践性很强。我想大致可将其分为“经院哲学”和“入世哲学”，马克思主义哲学属后者。它源于社会生活，富于严密分析，面向社会实践，是指导人类认识世界、改造社会的活动的哲学，世间少有。就社会生活和科学实践而言，唯物观和辩证思维是马克思主义哲学指导我们的核心思想。

在这样的背景下，我给博士研究生开设了“人工智能哲学”课程^①，无非想通过课堂来谈关于哲学及其与人工智能研究交汇方面的基本观点和核心思想，让同学们能在方法论角度得到启发，而不仅是学习具体的某种如“计算智能”方面的算法（在目前冯氏结构计算机上进行智能模拟，这是必需的）。博登的《人工智能哲学》（中文版由现在汕头大学工作的刘西瑞教授等翻译）一书包含了若干经典的人工智能哲学方面的文献，原文经典、精到译

^① 有同学上第一堂“人工智能哲学”课后问道，哲学是否讲世界观、方法论。我说，是的，只是我们开的不是“哲学”课，而是“人工智能哲学”课。



文。我们的课程离不开“弱人工智能与强人工智能”、“人工智能能否实现”等话题，但更注重以历史为鉴，分析存在的问题，并从有中国传统哲学思想特色的观点出发，探索人工智能的研究方法。

我先后尝试用我主讲、学生先在课前阅读讲课内容并在课堂提问等不同方式安排课程，同学们的提问、交流必然加深了我的认识，对内容的逐步完善做出贡献。例如，有同学问，禅宗是唯心主义还是唯物主义？回答当然是“唯心主义”。这使我进一步思考唯物主义和唯心主义各自的特点和意义。黑格尔的辩证法是唯心主义的，这不影响它在哲学史上的巨大价值。

应周昌乐教授之邀，我从2007年开始参加他和他的母校北京大学哲学系背景的同仁组织的“心灵与机器”年度研讨会，有幸向我国科学哲学界和人工智能界对人工智能哲学有研究或有兴趣的前辈、同道学习、讨论。这既使我们的“人工智能哲学”课没有成为“孤家寡人”，也使我有了更多学习的机会。曾有研究哲学的学者说，中国没有哲学家。我问：就当代而言，李泽厚先生算不算一个？回答说其“批判哲学之批判”是受了旁人启发而成的。我问：那个旁人是不是中国人？回答说：是的。事实和结论如何，由哲学界评判。但“中国没有哲学家”的说法显现出片面化的倾向，缺乏历史视野，带有局限性。讲点辩证法，恐怕就不至于此。

2001年10月，在潘云鹤老师办公室，我曾讲起“人工智能哲学”研究话题，潘老师认为宜做具体的形象思维方面的研究。这些年来，我个人本未在人工智能哲学方面有更多创见。与其说本书是关于人工智能哲学问题的概述，毋宁说是我对人工智能哲学的不完整的看法及由此所做的有限探索，难脱以偏概全和孤陋疏浅之责，仅以此抛砖引玉，恳请方家指正。

本书的出版获得“上海市优秀学科带头人计划”、“中国科学院百人计划”的资助。

目 录

前言

导论	1
----------	---

参考文献	4
------------	---

第一章 人工智能简史与现状	5
---------------------	---

一、实践和逻辑基础	6
-----------------	---

二、计算机器与智能	7
-----------------	---

三、主要观点	9
--------------	---

（一）符号主义	9
---------------	---

（二）联结主义	10
---------------	----

（三）行为主义	11
---------------	----

（四）单体与分布	12
----------------	----

四、专家系统	14
--------------	----

五、人工智能在中国	16
-----------------	----

六、困难	18
------------	----

小结	20
----------	----

参考文献	20
------------	----

进一步阅读资料	22
---------------	----

第二章 经典争论与基本问题	23
---------------------	----

一、相关观念	24
--------------	----

（一）唯物论和辩证法	24
------------------	----

（二）思维、认知和智能	26
-------------------	----

(三) 心智哲学与认知科学	29
二、图灵测试	31
三、塞尔中文屋	33
四、符号系统与神经网络	35
五、人工生命与行为主义	36
六、人类智慧与人工智能	38
七、计算的局限性	41
小结	43
参考文献	43
进一步阅读资料	46
第三章 抽象思维与形象思维	47
一、思维科学与智能工程	47
(一) 复杂系统	49
(二) 形象思维	50
(三) 人机结合	51
(四) 性智与量智	52
二、统计方法	54
三、面向对象与分析综合	56
四、图灵测试意义上的一种新形式	59
小结	65
参考文献	66
进一步阅读资料	67
第四章 中国文化与整体思维	68
一、禅宗思维	69
(一) 本体	69
(二) 禅宗简介	70
(三) 禅宗与艺术	72
二、灵感与顿悟	86

三、混沌动力学与悖论	88
(一) 混沌简介	88
(二) 悖论分析	89
(三) 悖论的动力学系统表示	94
(四) 本质与整体	95
小结	96
参考文献	96
进一步阅读资料	99
 第五章 智能模拟与经验知识	100
一、从智能模拟出发	100
二、经验	102
(一) 专家系统和语言能力	102
(二) 知识	104
(三) 逻辑与经验	105
三、计算机辅助医疗诊断	107
四、中医现代化	115
(一) 思想来源	118
(二) 系统思维	119
(三) 脉诊模拟	121
五、挑战	127
小结	131
参考文献	132
进一步阅读资料	135
 附录 A 计算机与智能	136
A1 模仿游戏	136
A2 新问题的评论	137
A3 参与游戏的机器	139
A4 数字计算机	140

◆ 人工智能哲学

A5 数字计算机的普适性	142
A6 对主要问题的反对意见	145
神学反对意见	146
“把头埋在沙中”的反对意见	147
数学上的反对意见	147
有关意识的论点	148
有关各种能力缺陷的论点	150
洛夫莱斯夫人的反对意见	152
有关神经系统连续性的论点	154
有关行为的非形式特性的特点	154
有关超感知觉的论点	155
A7 学习机	156
参考文献	162
附录 B 阴阳五行的整体思想	163
参考文献	171
附录 C 第 93 期东方科技论坛——智能模拟、思维模式和认知模型 简报	172
附录 D 第 134 期东方科技论坛——仿造脑及其应用问题简报	178
附录 E “心灵与机器” 2010 年苏州会议会后感	183

导 论

计算机问世后，人们就有可能开发出这样的系统，其复杂性层次是过去想都不敢想的；由于计算机能够计算和执行其内存的程序，启动了人工智能的研究^[1]。人工智能^①希望用计算机模拟智能过程和行为，并据此加深对人脑的功能和智能本质的认识和理解，但不限于生物学或神经科学可解释的意义。这样的企图至少蕴涵着三个方面的挑战：

第一，概念及其表达，诸如智能的本质，这些本质的特征，这些特征的形式化描述，智能与思维、认知、心智等的关系。

第二，人用计算机模拟其自身智能的可能性及其衡量原则，如弱人工智能与强人工智能、图灵测试和塞尔中文屋。

第三，进行探索的方法和可行性，什么样的案例可用以说明其意义和作用。

第一方面，显然也是哲学的基本内容；第二方面，一直是国际上人工智

^① 严格讲，人工智能是一个学科方向，不代表具体技术，用做主语似有欠妥之处。但通常已如此表述，此处姑妄从之。

能在哲学层面谈论的主题；第三方面，亦即方法论，是哲学的当然的研究目标。因此，人工智能与哲学有着天生的和不可分割的关系，人工智能和哲学两者在人工智能哲学的语境下比其他哲学分支更为互相包含、互相渗透，于是它是一个自然而然的话题。

人工智能哲学以人工智能研究中的哲学概念和争论为桥梁，讨论以计算机为载体的智能模拟、思维模式、认知模型的基本问题，包括模拟实现、模式刻画和模型构建的途径，并努力依托认知神经科学认识这些它们的物质基础，以系统视野和整体观点看待过去的曲折、分析目前的问题、探索有效的方法，指导人工智能的研究和实践。

拙著内容包括人工智能历史回顾、人工智能哲学的基本内容、思维科学与富有中国传统文化特色的禅宗思维、经验的重要意义以及如何获得有效的解决问题的方法等。

第一章是人工智能发展的历史回顾，后面的章节将在人工智能历史的基础上，讨论人工智能哲学的若干基本问题，强调系统观点和整体思维，进而讨论实际的知识获取、专家系统、知识工程和案例分析问题。

第二章在梳理人工智能的基本观念和思想以后，分析了人工智能哲学的若干基本问题和结论。还原论思想既没有提供解决根本问题的手段，也不足以给予宏观指导；基于冯氏结构计算机的局限性是明显的。指出抽象思维与形象思维结合、人与机器交互是合理之路。

第三章首先介绍钱学森先生倡导的包括形象思维和抽象思维的思维科学的核心观点，这是有中国特色的人工智能哲学思想。进而提出统计方法能同时体现形象思维和抽象思维的特征，并阐述面向对象这一企图模拟人的思维特点的方法的哲学内涵，还介绍了体现抽象思维与形象思维两者的书法创作模拟这一图灵测试意义下的新形式。

第四章讨论禅宗的整体思维特征，这是第三章介绍的思维科学体系涉及顿悟概念的具体化，极具中国特色。这里还以书法艺术为例说明禅宗思维与艺术创作过程互相交融和影响的呈现形式和独特之处。用混沌动力学系统刻画、体现整体思维的尝试也在此做了简单介绍。

第五章刻画从智能模拟、思维模式到认知模型几个阶段的各自内涵和相

互关系，对前面几章思想的总结，又以计算机辅助医疗诊断、包括中医现代化为例，对实际应用和面临挑战给出示例，说明人工智能哲学的指导意义。这里强调经验知识在人工智能研究中的根本作用。

历史的回顾有待充实，哲学问题的讨论不够全面，思维科学并不在整个人工智能界为大多数学者所理解或接受，禅宗思维更是有“唯心”之嫌。但重理论联系实际、强调经验知识是基本要旨。

概言之，拙著涉及的人工智能哲学的基本内容及其关系如下。

$PAI \triangleq \{O, M, P, F, B\};$

//人工智能哲学涉及人工智能的目标(Object)、方法论(Methodology)、进程(Progress)、原则(Fundamental)和瓶颈(Bottleneck)问题。

$Object \triangleq IS \cap IU;$

//目标是模拟(Simulation)智能和理解(Understanding)智能。

$I \triangleq \{T, \dots\};$

//智能模拟的核心是思维(Thinking)模拟。

$T \triangleq TA \cap TI;$

//思维包括抽象(Abstract)思维和形象(Imagery)思维。

$M \triangleq S(T, H \cup C);$

//方法论是整体化(Synthesis),即抽象思维和形象思维的结合,人(Human)和机器(Computer)的交互。

$P \triangleq \{IS, TP, CM\};$

$ISC \cap TP \subset CM;$

//进程是从智能模拟(Simulation)经思维模式(Patterns)到认知模型(Cognition Model)的递进。

//相应外延逐步扩大。

$F \triangleq Ma \cap Di;$

//原则是唯物论(Materialism)和辩证法(Dialectic)。

$B \triangleq \{L, K\};$

//瓶颈是学习(Learn)和知识(Knowledge)。

$K \triangleq \{Op, On\};$

$K \triangleq \{Op, On\};$

//有关知识的主要方面是操作(Operation)和内容(Content)。

$Op \triangleq \{A, E, R\};$

//操作包括获取(Acquisition)、表达(Express)和推理(Reasoning)。

$Co \triangleq \{F, C\};$

//内容包括领域(Field)知识和常(Common)识。

$F \triangleq \{Ex, Im\};$

//领域知识包括显式(Explicit)知识和隐性(Implicit)知识。隐性(Implicit)知识主要与经验有关。

拙著是在这个框架下的尝试,有待不断深化和完善。

参 考 文 献

- 1 [美] 司马贺. 人工科学: 复杂性面面观. 武夷山译. 上海: 上海科学技术出版社, 2004: 160

第一章 人工智能简史与现状

……尽管历史充满了盛衰浮沉，但却存在一种总的趋势，我们从中仍然可以感到某种满意：我们比我们的祖先知道得更多，我们更能够驾驭自然界，我们所遭受的疾病和自然灾害要更少一些。……

对历史的透视能够使我们更清楚地看出，什么事件和哪种活动有着永久的重要性。大多数与伽利略同时代的人，把30年战争看得比他的发现更为重要；但是对于我们来说，这次战争是30年光阴的虚掷，而他的发现却开辟了一个新时代。当格拉斯顿拜访达尔文后，达尔文说：受到这样一位伟大人物的拜访，是多么的荣耀。他的谦逊是可爱的，不过也表明了他缺乏一种历史眼光。许多事情——如党派斗争——可以在当时激起与它们真正的重要性十分不成比例的事情，而最伟大的事件则有如高山的顶峰，尽管是雄踞在远处，却被视界较近的前景遮蔽了。要学会从它们的历史背景去观看当代的事情，并把它们想象成过去所呈现的样子——这种习惯会有助于健全而冷静的判断^[1]。

——罗素

迄今为止,人工智能尽管能战胜国际象棋冠军,但下围棋时无能为力;虽然能读文章,但创作时十分吃力;已能进行一定程度的学习,但没有通用的应用方式,不能识别任意角度的对象,不能在摔跤时拣起有用的东西……甚至还不具备人类的一些简单的思维功能,如婴儿的形象思维能力。人工智能距其初衷和目标依然很遥远。为了了解人工智能发展的曲折过程并以史为鉴,本章简要回顾人工智能发展的历史和其间出现的问题。

一、实践和逻辑基础

人类迈向人工智能的脚步可上溯到西方亚里士多德(Aristotle)时期和我国先秦诸子百家时代的形式逻辑,前者包括演绎推理的三段论,后者如公孙龙《白马论》中的“白马非马”辩论、《列子·汤问》中关于几乎能乱真的机械人的描述。

17世纪,帕斯卡(Blaise Pascal)和莱布尼茨(Wilhelm Leibniz)萌生了有智能的机器的想法,帕斯卡制造了“加法器”。1822年,巴比奇(Charles Babbage)制造出了“差分机”,这台基于有限差分法的机器能够计算数学用表的内容,并且把计算结果一行一行打印出来,首先被应用于航海。这台机器具有独特的机械结构,因此还不能和电子计算机简单直接地进行比较^[2]。巴比奇设计制造“差分机”、“分析机”最终以失败告终。19世纪末至20世纪初,罗素(Bertrand Russell)、怀特海(Alfred North Whitehead)和哥德尔(Kurt Gödel)集大成并对之做出杰出贡献的数理逻辑,是经典人工智能的理论基础。哥德尔与亚里士多德、莱布尼茨均为逻辑学历史上最杰出的人物,罗素的工作是其研究参考点。1934年,哥德尔引进了一般递归函数概念,后来发现它也捕获了直观的机械程序概念。他认为,只有图灵(Alan Mathison Turing)1936年发表的《论可计算数及其在判定问题中的应用》才第一次摆出令人信服的分析,告诉我们怎样用正确的眼光来看清直观概念。王浩曾向哥德尔询问,如果允许一定的误差概率的话,是否每台模拟计算机都能够被一台数字计算机完全恰当地临摹出来^[2],但哥德尔对计算机的实际发展没有兴趣。



二、计算机器与智能

人工智能最早的奠基工作是由英国逻辑学和计算机的先驱者图灵开展的^[3]，他在《论可计算及其在判定问题中的应用》论文中阐明了现代计算与计算机原理，设想出抽象的数字计算机，美国逻辑学家丘奇（Alonzo Church）把它叫做“图灵机”。图灵指出把基于离散量的递归函数作为智能描述的基础，证明了必然存在无限多个不可计算的问题，即存在无限多个问题无法用计算机程序来解决；还提出了一个计算机不能解决的问题，即停机问题。

丘奇、图灵等对计算本质的思考，如计算和符号处理，数字并非计算的主要方面等^[4]观点，是同人工智能的基本思想相一致的。他们还提出了著名的丘奇-图灵论题，即凡是可计算的函数都是一般递归函数，都可以用一台图灵机来计算。

图灵在1936年的工作中还指出了自指逻辑难题，哥德尔从康托尔（Georg Cantor）和罗素的工作中使其得到了进一步的发展，在十年的时间里图灵把通用机理论转变为现代意义上的计算机的实际发明。图灵早在1941年就开始思考智能机器，特别是那些由经验法则（后被称做“启发式搜索”）来指导、通过搜索可能的解空间求解问题的可能性，还有国际象棋的机械化等。“启发式搜索”思想起源于图灵设计的用于破解密码的机器智能历史上的第一个里程碑“Bombe”。图灵还试验了两种后来在人工智能中通用的启发式搜索规则：“最小最大化”（Minimax，假设对手会以收益最大化的方式移动棋子，通过使对手期望的移动导致自己的损失最小化来移动自己的棋子）以及“最好优先”（Best first，按一个数字化的记分系统列出所有可用的移动，首先检查最高分的移动的结果）。

相对计算的实际应用：图灵对产生脑活动模型的可能性更感兴趣。他认为人工智能的目的是揭示心智或者智能是如何工作的，并总结了人工智能和认知科学之间的联系。他指出：完整的思考过程仍然是非常神秘的，但是试着建造一台思维机器非常有助于弄清楚我们是如何思考的。